Inference in nonlinear systems with unscented Kalman filters

Diana Giurghita, *University of Glasgow*, D.Giurghita.1@research.gla.ac.uk
Dirk Husmeier, *University of Glasgow*, Dirk.Husmeier@glasgow.ac.uk

Abstract. An increasing number of scientific disciplines, most notably the life sciences and health care, have become more quantitative, describing complex systems with coupled nonlinear differential equations. While powerful algorithms for numerical simulations from these systems have been developed, statistical inference of the system parameters is still a challenging problem. A promising approach is based on the unscented Kalman filter (UKF), which has seen a variety of recent applications, from soft tissue mechanics to chemical kinetics. The present study investigates the dependence of the accuracy of parameter estimation on the initialisation. Based on three toy systems that capture typical features of real-world complex systems: limit cycles, chaotic attractors and intrinsic stochasticity, we carry out repeated simulations on a large range of independent data instantiations. Our study allows a quantification of the accuracy of inference, measured in terms of two alternative distance measures in function and parameter space, in dependence on the initial deviation from the ground truth.

Keywords. Unscented Kalman Filter, Parameter estimation, Nonlinear models, Ordinary differential equations, Stochastic differential equations.

1 Introduction

Mathematics is transforming biology in the same way it shaped physics in the previous centuries [2]. The underlying paradigm shift that distinguishes modern quantitative systems biology from more traditional non-quantitative approaches is based on a conceptualisation of elementary processes as a complex network of interactions, and its representation with an adequate mathematical description, typically in terms of coupled differential equations. While the intrinsic nonlinearities typically defy analytical tractability, advances in high-performance computing provide the hardware for fast numerical solutions. This allows an in silico exploration of complex biological systems under varying experimental conditions and in different environmental contexts. However, this forward modelling approach (simulation via numerical solution of a given mathematical description) assumes that the system under investigation is known, i.e. that the parameters defining the kinetics and dynamics of the interactions are given. Such detailed

knowledge is rarely available in practice. What is needed is a solution of the so-called inverse problem, i.e. the rigour of statistical inference to systematically infer the kinetic parameters from given data.

The direct approach to parameter estimation is to minimise a divergence measure between the predictions from the model and the data, like in [3]. This approach is computationally expensive and suffers from susceptibility to local optima. Gradient matching bypasses the computationally expensive numerical solution of the differential equations and thereby allows a more exhaustive exploration of the parameter space. However, this comes at the price of information loss inherent in gradient matching, which is the subject of current methodological research [5]. A promising idea is based on Bayesian filtering, and in particular the unscented Kalman filter (UKF). The idea of the UKF, as proposed in [7], is to relax the linearity constraint of the underlying dynamics without incurring an explosion of the computational complexity (as opposed e.g. to particle filters [6]), and to include the unknown system parameters in an augmented state vector, so as to automatically track them with established Bayesian filtering techniques. In more detail, the mathematical description of the biological system leads to iterative nonlinear equations relating an augmented state vector at the present time point to a nonlinear function of the state vector at the previous time point perturbed by additive noise. The noise distribution is assumed to be multivariate normal. Starting from an initial distribution of state vectors, drawn from a multivariate normal distribution with an initial covariance matrix, state vectors are iteratively subjected to the nonlinear dynamics of the state equations, which can easily be parallelized. From these sampled vectors, the so-called sigma points, the new covariance matrix is computed. By iterative assimilation of new measurements and application of established and computationally efficient Kalman filtering techniques [6], we can obtain the posterior distribution of the model parameters. This procedure was successfully applied to inference in soft tissue mechanics of the heart [9] and chemical kinetics [1].

A potential drawback of the UKF is the dependence of the posterior distribution on the initial state, as pointed out in [8]. Given the Markovian nature of the process, the initialisation should not matter if the system is ergodic, but there is no guarantee that this condition is met in practice. The performance presented in the seminal article of [7] is impressive, but a closer inspection reveals that all results were obtained from highly informative initialisations, which started from values close to the true parameters. The objective of the present article is to investigate the dependence on the initialisation more systematically, based on an extensive range of numerical simulations. To this end, we choose the systems in [7], which are representative of what we typically encounter in systems biology: (1) a deterministic dynamical system with periodic attractor, (2) a deterministic dynamical system with chaotic attractor, and (3) a stochastic system. We systematically quantify the accuracy of parameter inference based on two alternative divergence measures: mean square error in function space, and relative bias in parameter space. Our study provides practical guidelines about the robustness of inference with UKF, and indications of when complementary techniques for informed initialisation, e.g. [8], are needed.

2 Methods

This section provides a short summary of important UKF concepts, as provided in [6], where the reader will find more details, discussion and derivations on the topic. The general form of the

state space model consists of the following state and observation equations: $\mathbf{x}_t = g(\mathbf{x}_{t-1}, \mathbf{u}_t, \boldsymbol{\epsilon}_t)$, $\mathbf{y}_t = h(\mathbf{x}_{t-1}, \mathbf{u}_t, \boldsymbol{\eta}_t)$, where $\mathbf{x}_t$ is the hidden state, $\mathbf{u}_t$ is an optional input or control signal, $\mathbf{y}_t$ is the observation, g is the transition model describing the system dynamics, h is the observation model, and $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ representing the system respectively observation noises at time t. The Kalman filter algorithm performs exact Bayesian filtering for linear-Gaussian state space models (where g and h are linear and $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are Gaussian). The advantage of the method comes from the fact that the probability distribution of the predictor step: $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and the probability distribution of the updating step $p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{y}_{1:t-1})$ can be obtained in closed form because of the Gaussian assumption (see [6] for derivations).

However, in practice, models are often nonlinear and the Gaussian assumption does not hold and the UKF is one of the methods that can accommodate such scenarios. The UKF, proposed by Julier and Uhlmann [4] is based on the idea that it is easier to approximate a Gaussian than to linearise a function. First, a set of points, called sigma points, are chosen in a deterministic way, and then passed through a nonlinear function. Second, a Gaussian distribution is fitted to the resulting transformed sigma points. This is called the unscented transform and it is performed as follows. Given $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, suppose we want to estimate $p(\mathbf{y})$ where $\mathbf{y} = \mathbf{f}(\mathbf{x})$ and $\mathbf{f}$ is a nonlinear function. A set of $2d + 1$ deterministically chosen points, the sigma-points are obtained as follows: $\mathbf{x} = (\boldsymbol{\mu}, \{\boldsymbol{\mu} + (\sqrt{(d+\lambda)\boldsymbol{\Sigma}})_{:i}\}_{i=1}^{d}, \{\boldsymbol{\mu} - (\sqrt{(d+\lambda)\boldsymbol{\Sigma}})_{:i}\}_{i=1}^{d})$, where λ is a scaling parameter, d is the dimension of $\mathbf{x}$, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}_{:i}$ represents the i'th column of the covariance matrix. The sigma points are propagated through f to obtain $\mathbf{y}_i$, and the mean and covariance of $\mathbf{y}$ are calculated based on the weighted $\mathbf{y}_i$'s. See [6] for more details about the unscented transform. The UKF uses the unscented transform twice. First, the sigma points are transformed using the system dynamics model g to compute the prediction distribution: $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and, second, the sigma points are transformed using the measurement model function h to compute the update distribution: $p(\mathbf{x}_t|\mathbf{y}_{1:t})$.

Parameter estimation can be accomplished with the UKF [7] by considering the parameter vector $\boldsymbol{\lambda}$ as a dynamical variable that follows the dynamical model: $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_{t-1}$. Similarly, the process noise, $\boldsymbol{\epsilon}_t$, and the measurement noise, $\boldsymbol{\eta}_t$, are added to the state vector, resulting in the joint state vector $\mathbf{j}_t$ that has the following state and observation equations:

$$\mathbf{j}_t = \begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\lambda}_t \\ \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{bmatrix} = \begin{bmatrix} \mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\lambda}_{t-1}) + \boldsymbol{\epsilon}_{t-1} \\ \boldsymbol{\lambda}_{t-1} \\ \boldsymbol{\epsilon}_{t-1} \\ \boldsymbol{\eta}_{t-1} \end{bmatrix} = \mathbf{g}^j(\mathbf{j}_{t-1}), \qquad \mathbf{y}_t = \mathbf{h}^j(\mathbf{j}_t) = \mathbf{h}(\mathbf{x}_t) + \mathbf{h}^\eta(\boldsymbol{\eta}_t) \tag{1}$$

In the models considered in this paper $\mathbf{h}^\eta$ is the identity matrix, but in scenarios with correlated noise it will be a non-diagonal matrix.

3 Data

This section presents the three models used in [7]: the Lotka-Volterra system, the chaotic Lorenz system, the stochastic van der Pol system, and the results of the augmented UKF (as described in Section 2) for signal tracking and parameter estimation.

Lotka-Volterra system

The Lotka-Volterra (LV) system is structured as a system of two ordinary differential equations (ODEs) as follows:

$$\frac{dx_{1t}}{dt} = \lambda_{1t}x_{1t} - \lambda_{2t}x_{1t}x_{2t} \qquad \frac{dx_{2t}}{dt} = \lambda_{2t}x_{1t}x_{2t} - \lambda_{3t}x_{2t} \tag{2}$$

The model describes the interaction of prey and predator populations which are represented as concentrations by the variables (x_2) respectively (x_1). The parameters: $\lambda_{1t}, \lambda_{2t}, \lambda_{3t}$ representing the growth rate of prey population, death rate of predator population, respectively growth rate of the prey population are considered unknown and augmented to the state variables for the joint estimation using the UKF, as discussed in Section 2 . Data was generated by numerical integration of the equations in (2), using a Runge-Kutta method implemented in `Matlab` by function `ode45`. The sampling step size is $\Delta t = 0.1$, the parameters are $\lambda_{1t} = 1, \lambda_{2t} = 1.5, \lambda_{3t} = 2$ and the initial values for the numerical integration are $\mathbf{x}_0 = (0.5, 1)^T$. To obtain the measurements, the concentration of prey population x_{1t} has been corrupted with additive Gaussian noise with standard deviation 0.1: $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{R})$ to ensure a signal to noise ratio (SNR) of 8.26. Thus,

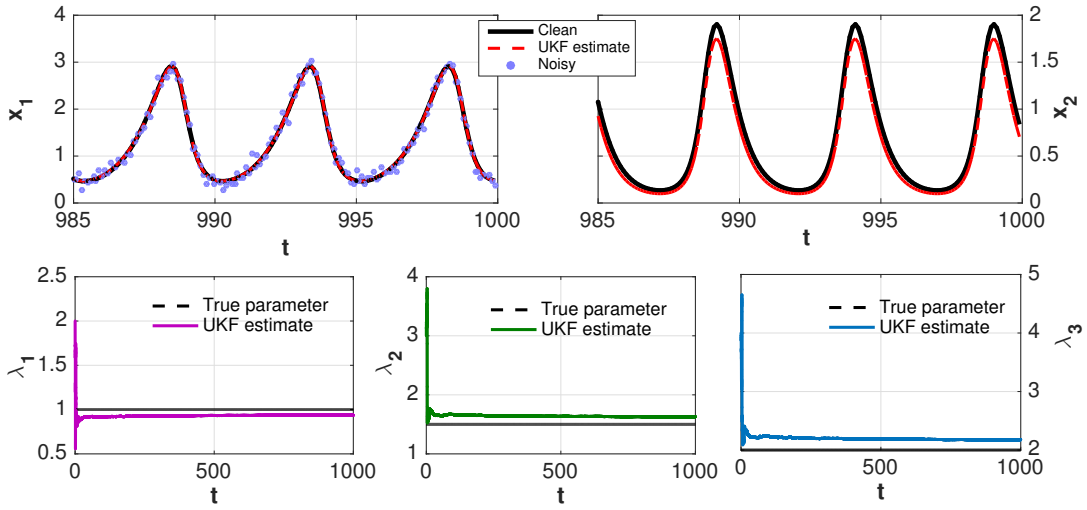

Figure 1. **Lotka-Volterra model** with parameters $\lambda_1 = 1, \lambda_2 = 1.5, \lambda_3 = 2$. *Top*: UKF tracking of prey and predator population concentrations for $t \in [998, 1000]$. *Bottom*: UKF estimation for Lotka-Volterra parameters.

the joint state space vector for the UKF contains the population densities $\mathbf{x}_t = (x_{1t}, x_{2t})^T$, the unknown parameters $\boldsymbol{\lambda} = (\lambda_{1t}, \lambda_{2t}, \lambda_{3t})^T$, as well as the measurement noise $\boldsymbol{\eta}_t$ which is assumed to be known, as detailed in Section 2. The initialisation for the UKF is the first observation for the states: $\hat{\mathbf{x}}_0 = (y_0, y_0)^T$, and for the parameters twice the true values: $\hat{\boldsymbol{\lambda}}_0 = (2, 3, 4)$. The covariance matrix is initialised as the identity matrix: $\hat{\mathbf{P}}_0 = \mathbb{I}_5$.

The resulting limit cycle for the chosen set of parameters displays periodic oscillations which can be observed in Figure 1. The UKF algorithm tracks the real signal closely, with the estimates overlapping the clean solution towards the end of the observation time window. This is due to the fact that the UKF initialisation at y_0 is very close to the actual value given the small amount

of measurement noise in the data. Furthermore, the unobserved state x_2, and the parameters are estimated with high precision using the UKF, as suggested by the small standard errors reported in Table 1.

Lorenz system

The Lorenz system has been extensively studied in relation to models of Earth's atmospheric convection. The system's chaotic behaviour poses an interesting challenge for the UKF in terms of convergence, especially since the system dynamics can drastically change depending on initial conditions. The mathematical form of the Lorenz system consists of the following three ODEs:

$$\frac{dx_{1t}}{dt} = -\lambda_{1t}x_{1t} + \lambda_{1t}x_{2t}; \quad \frac{dx_{2t}}{dt} = \lambda_{2t}x_{1t} - x_{2t} - x_{1t}x_{3t}; \quad \frac{dx_{3t}}{dt} = -\lambda_{3t}x_{3t} + x_{1t}x_{2t} \tag{3}$$

The ODEs in (3) with parameters $\lambda_{1t} = 10, \lambda_{2t} = 46, \lambda_{3t} = 8/3$, and initial conditions $\mathbf{x}_0 = (1, 1, 1)$ are numerically integrated using a Runge-Kutta method. $N = 10000$ data samples are obtained at equidistant intervals, $\Delta t = 0.01$, and the measurements are obtained by adding Gaussian noise $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ to the x_{1t} component only. To reproduce the results in Sitz et al. [7], the noise variance is chosen as: $\mathbf{R} = 4$. The initialisation for the UKF is the first observation for the states: $\hat{\mathbf{x}}_0 = (y_0, y_0, y_0)^T$ and for the parameters it is half the true values: $\hat{\boldsymbol{\lambda}}_0 = (5, 23, 4/3)^T$. The covariance matrix is initialised as a diagonal matrix: $\hat{\mathbf{P}}_0 = 10\mathbb{I}_6$.

In Figure 2(*Left*), the complexity of the system dynamics can be recognised by looking at the nonlinearities in 2D projections and non-periodic oscillations of the $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ components in time. The UKF proves it is able to reconstruct the original signal (Figure 2, *Right*) providing highly accurate parameter tracking with bias within three decimal points and standard errors less than 1%, as reported in Table 1. As before, this is a consequence of the initialisation for the observed state that is very close to the true values due to the low level of measurement noise.

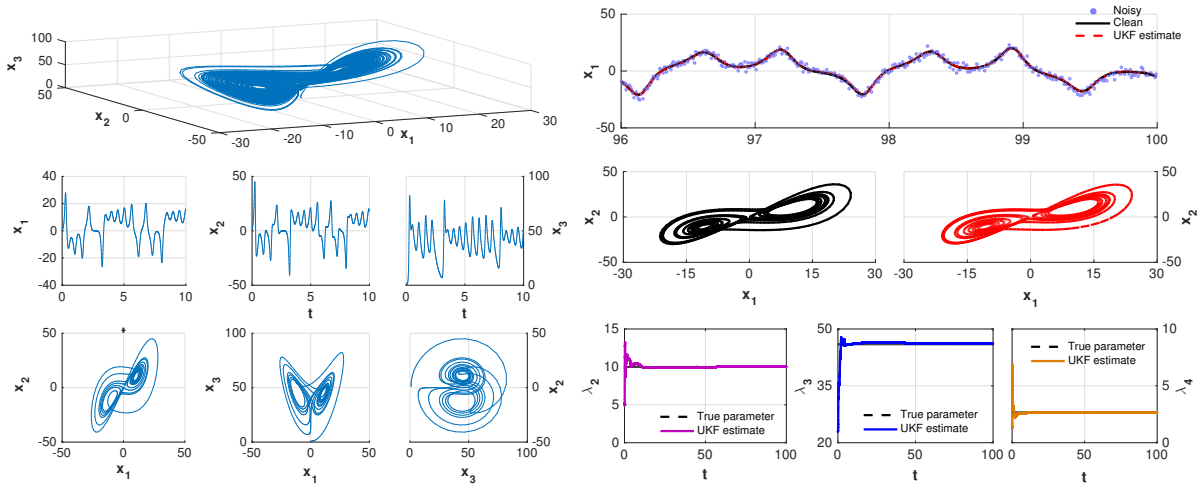

Figure 2. **Lorenz attractor** with parameters $\lambda_1 = 10, \lambda_2 = 46, \lambda_3 = 8/3$. *Left:* System dynamics and projections in 2D and 1D. *Right:* Signal reconstruction for $t = [98, 100]$, projection on 2D using actual and UKF estimated components x_1 and x_2, and parameter estimation using the UKF.

Van der Pol system

Consider the following equations describing the van der Pol oscillator:

$$\frac{dx_{1t}}{dt} = x_{2t}, \quad \frac{dx_{2t}}{dt} = \lambda_{1t}(1 - x_{1t}^2)x_{2t} - x_{1t} + \epsilon_t \tag{4}$$

The stochasticity of the system comes from the second component which contains an uncorrelated noise term ϵ_t. For the augmented UKF, the process noise is considered fixed and known: $\epsilon \sim N(0, 1)$. The noise propagates through both states x_1 and x_2, which means the system will display stochastic oscillations throughout the observation time window. See Figure 3 for a solution of the stochastic van der Pol system, with parameter $\lambda_{1t} = 3$ and $[1, 0]$ as the initial conditions for the numerical integration. Measurements are obtained from the first component, $\mathbf{x}_{1t}$ by adding Gaussian noise: $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ with variance $R = 0.2^2$. This ensures a SNR of around 10%. As with all stochastic systems, the numerical integration has to be performed using a very small step size, $\delta t = 0.001$, and a low order integration method such as the Euler method [7].

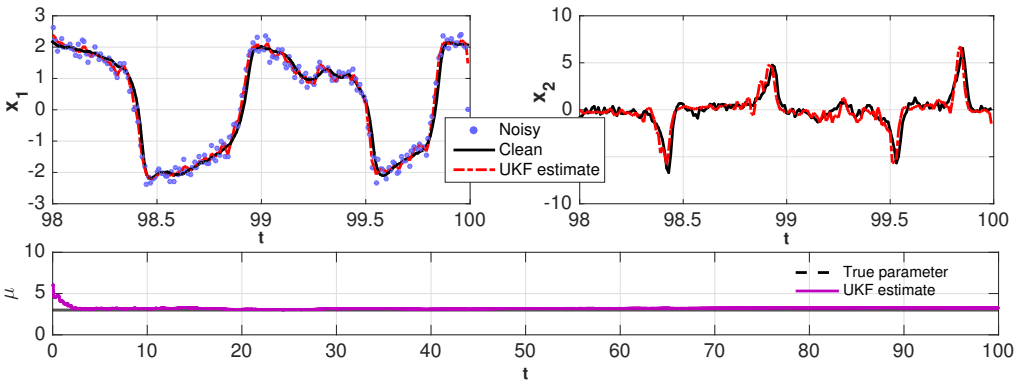

Figure 3. **Stochastic van der Pol system** with parameter $\lambda_1 = 3$: signal reconstruction (*Top*) and parameter estimation (*Bottom*) using the UKF.

The initialisation for the UKF is the first observation for the states: $\hat{\mathbf{x}}_0 = (y_0, y_0)^T$, and for the parameter twice the true value: $\hat{\boldsymbol{\lambda}}_1 = 6$. The covariance matrix is initialised as a diagonal matrix: $\hat{\mathbf{P}}_0 = 2\mathbb{I}_3$. In Figure 3 and Table 1 we report the results of the UKF estimation for the van der Pol system. Since the UKF initialisation for the observed system state is relatively close to the true one (SNR is 10%), the estimated path of the signal is very close to the real signal. However, the tracking for the second component is not as precise, which is also reflected in the higher standard error for $\mathbf{x}_2$. Nevertheless, this is not unusual given the stochastic nature of the system, and the fact that this state is not observed.

To conclude this section, we report that our results from the UKF estimation for the three considered systems are consistent with those reported in [7].

4 Simulations

The main aim of this study is to systematically investigate the influence of the UKF initialisation, given that this is an important practical aspect that has not been addressed in the literature

Table 1. UKF estimation results for the Lotka-Volterra, Lorenz and van der Pol systems. Point estimates and standard errors (in brackets) are reported using the last prediction step of UKF.

	States and parameter estimates using the UKF					
	$\hat{\mathbf{x}}_1$	$\hat{\mathbf{x}}_2$	$\hat{\mathbf{x}}_3$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
Lotka-Volterra	0.474 (0.001)	0.837 (0.004)	-	1 (0.001)	1.5 (0.003)	2 (0.004)
Lorenz	-4.836(0.855)	-10.518(1.808)	17.665(0.635)	10(0.019)	46 (0.101)	2.667 (0.009)
Van der Pol	2.075 (0.111)	0.157 (0.627)	-	3 (0.093)	-	-

in great depth. We complement the results presented in [7] with a series of simulation studies to provide a more complete and informative picture of UKF estimation for nonlinear models. Hence, we preserve the modelling choices reported in [7]: sampling step size, integration step size, sampling time interval, system parameters, model errors etc. Note that the covariance matrix initialisation for the UKF will most likely be different from the one chosen by Sitz et al. [7], as it was not reported by the authors.

Initialisation

The impact of the initialisation has been assessed by considering a range of offsets for the states and model parameters: $\{0\%, 25\%, 50\%, 100\%, 150\%\}$, which include the choices in [7] across the 3 models included. For consistency, the offsets have been calculated as percentages from the true values, with the choice of a positive or negative offset being decided at random for each UKF instantiation. The offset has been considered for the states $\mathbf{x}_0$ as well, even though the most common scenario would be to initialise using the first observed states $\mathbf{y}_0$. The reasoning behind this choice is that, in practice, it is also plausible to consider a different initialisation other than the first observation, which would be based on prior knowledge of the system dynamics.

Number of observations

A secondary aim of this study is to investigate the impact of reducing the number of observations, given the results reported in Section 3 indicate that comparable UKF estimates could be obtained using a smaller number of observed samples. reaches Two scenarios have been compared, see Table 2: the 'Same frequency' scenario which relies on fixing the same sampling step size, Δt, and reducing the time window, T, to acquire the desired number of observations, N, and the 'Less frequency' scenario which relies on increasing the sampling step size until a set number of observations is achieved in a certain time window. Given that originally the results have been produced using $N = 10000$ observation, which is an excessive number for most applications in practice, we opted for a more realistic $N = 3000$ observations, which is a 70% decrease in the number of observations. These original settings, denoted as 'Original' in Table 2 have been used as a benchmark for the other two scenarios.

Reporting results

Two measures have been calculated for each scenario to asses differences in the performance in parameter space and in function space: relative bias $RelBias_t = \frac{\hat{\mathbf{x}}_t - \mathbf{x}_t}{\mathbf{x}_t}$ and mean squared error $MSE = \sum_{t=\frac{N}{2}}^{N} (\hat{\mathbf{x}}_t - \mathbf{x}_t)^2$. For the calculation of the MSE, the first half of the UKF estimation

Table 2. Details on the simulation scenarios for investigating the impact of sampling frequency on the UKF results.

	Lotka-Volterra system	Lorenz system	Van der Pol system
Original	$\Delta t = 0.1, N = 10000$ $T = [0, 1000]$	$\Delta t = 0.01, N = 10000$ $T = [0, 100]$	$\Delta t = 0.01, N = 10000$ $T = [0, 1000]$
Same frequency	$\Delta t = 0.1, N = 3000$ $T = [0, 300]$	$\Delta t = 0.01, N = 3000$ $T = [0, 30]$	$\Delta t = 0.1, N = 3000$ $T = [0, 300]$
Less frequency	$\Delta t = 0.33, N = 3000$ $T = [0, 1000]$	$\Delta t = 0.033, N = 3000$ $T = [0, 100]$	$\Delta t = 0.33, N = 3000,$ $T = [0, 1000]$

is discarded across the simulations, to ensure the UKF estimation has enough time to converge to the true value, which will also ensure a fairer comparison across data sets with very different initialisations. Although in recursive methods such as the UKF, the last step (after all the data has been seen) provides the best estimate for the state, in stochastic models the bias will often exhibit small fluctuations, in which case the bias reported has been calculated as an average of the biases for the last 100 estimates.

5 Results

This section contains the results for the simulations carried out to investigate the effect of the initialisation on parameter estimation using the augmented UKF, as well as the effect of reducing the sample size, as discussed in Section 4 and summarized in Table 2. Across the scenarios considered, 10 simulation have been run for each situation e.g.: 10 sets of 'Original', 10 sets of 'Same frequency', 10 set of 'Less frequency' etc. Without loss of generality, the results for the unobserved states have been excluded from the plots since they exhibit similar patterns as the observed system state.

Lotka Volterra system

The simulation results in Figure 4 suggest that decreasing the observed sampling time either by increasing the sampling frequency time step, or by reducing the time interval for the system measurements will result in a deterioration of the results, in terms of bias and MSE. Although the range of the bias and MSE suggests keeping the sampling step size constant, the median indicates that increasing the sampling size will increase the bias by on average 5%.

Figure 5 shows the effect of the initialisation offset on the UKF estimation for the three models considered, by looking at box plots of the relative bias from 10 data sets. For the Lotka Volterra model, as the relative offset increases to 150% the estimation deteriorates by, on average, 7 % for the parameters bias, and 0.7 % for the state bias. Notice that the bias range increases by a factor of between 4 and 7 for the parameters, which means the estimation becomes considerably uncertain for 150% offset. The MSE at 150% offset is, on average, about 50 times higher for signal and up to 10 times higher for parameters compared to 0% offset, indicating the UKF estimation is considerably worse at higher offset.

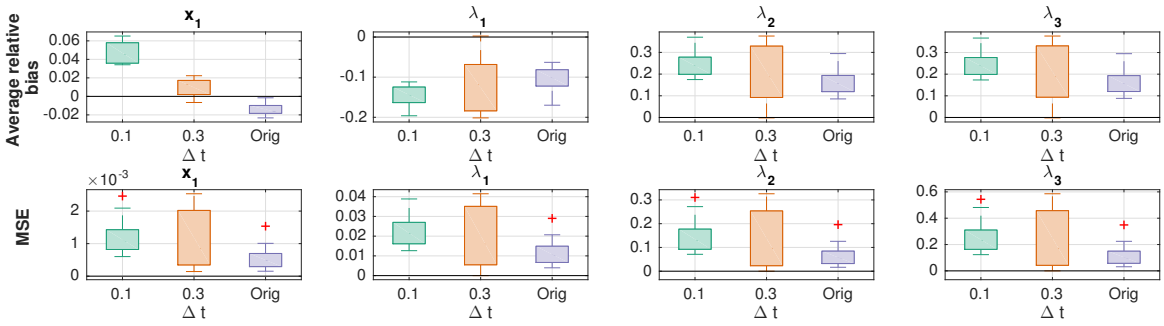

Figure 4. **Lotka-Volterra system**: observation sample size. Boxplots of relative bias (*Top*) and MSE (*Bottom*) over 10 data sets for the three scenarios considered in Table 2.

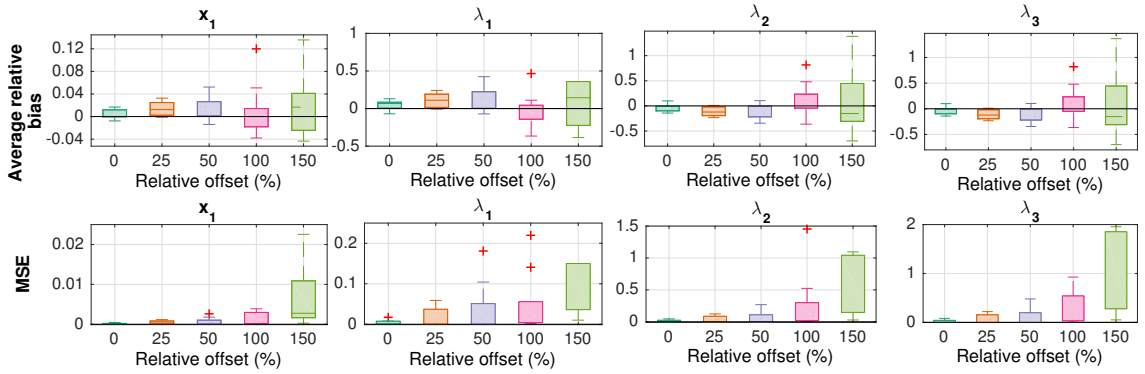

Figure 5. **Lotka-Volterra system**: initialisation. Boxplots of relative bias (*Top*) and MSE (*Bottom*) over 10 data sets with relative offsets:$\{0\%, 25\%, 50\%, 100\%, 150\%\}$.

Lorenz system

In Figure 6, the bias indicates that maintaining the sampling size at $\Delta t = 0.01$ produces slightly better results compared to decreasing the sampling frequency, although a difference of around 1-2% in parameter space is likely to be irrelevant in practice. However, the range of the bias and MSE suggests that UKF is more likely to produce estimates that are more inaccurate if the number of observations is reduced (by either scenario). Furthermore, the presence of outliers in reduced sample size scenarios means that the algorithm can get stuck in local optima of the likelihood landscape, and one way to overcome this would be a larger sample size. This is supported by the MSE which is, on average, up to 10 times higher for parameter estimates for two scenarios compared to the 'Original' scenario.

Figure 7 shows results for only 4 of the 5 offsets considered in the other scenarios due to the fact that in the chaotic Lorenz system a large initial offset is very likely to produce very different system dynamics. For example, with an initial offset of 150% the UKF performed very poorly due to numerical instabilities which caused the algorithm to stop before the end, as such results are not available for the comparison. Note also that the estimation degrades for the parameter bias, on average, up to 50% and up to a factor of 150 for the parameter MSE at 100% offset.

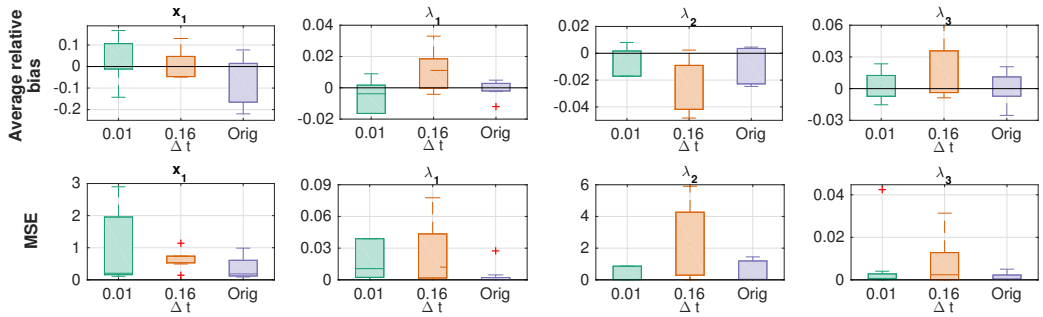

Figure 6. **Lorenz system**: observation sample size. Boxplots of relative bias (*Top*) and MSE (*Bottom*) over 10 data sets for the three scenarios considered in Table 2.

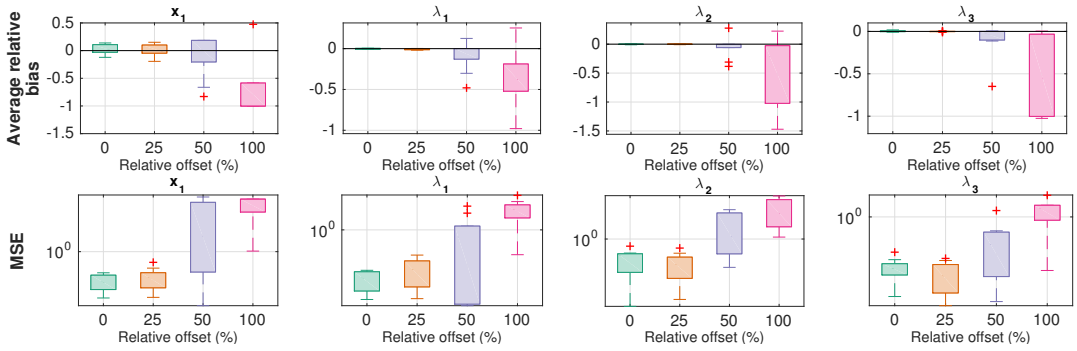

Figure 7. **Lorenz system**: initialisation. Boxplots of relative bias, zoomed-in to a more informative scale (*Top*), and MSE (*Bottom*), on a log scale due to high range of values. Estimates obtained over 10 data sets with relative offsets:$\{0\%, 25\%, 50\%, 100\%\}$.

Van der Pol system

In Figure 8, the relative bias and the MSE indicate that reducing the sampling frequency reduces the quality of the estimation in terms of parameter bias by, on average, 140%, and in terms of MSE by a factor of 2 on the $\log_{10}$ scale. The stochasticity of the system poses a greater challenge for the UKF, and the increase in sampling frequency is leading to a significant loss of information that is required to provide more accurate estimates. Given that the difference between the bias of the 'Original' and '$\Delta t = 0.1$', is less than 1% means the UKF achieves convergence really fast, even before the first 30% of the observations. This makes sense given the periodic oscillation in the observed system. When dealing with similar periodic systems, having data that covers several cycles of the observed signal would be sufficient for the UKF to provide reliable estimates. In Figure 9 we present the results of UKF estimation of the van der Pol system, with different initialisations. UKF estimation is quite robust to initialization for the van der Pol system, even for 150% offset, with the results indicating the bias for all states is, on average, close to 0 for all initialisations. For the parameter λ_1 the estimation differs only by 1 to 2% on average. The MSE is similar in terms of median and range across offsets, indicating that reliable estimates can be obtained from a big range of starting points.

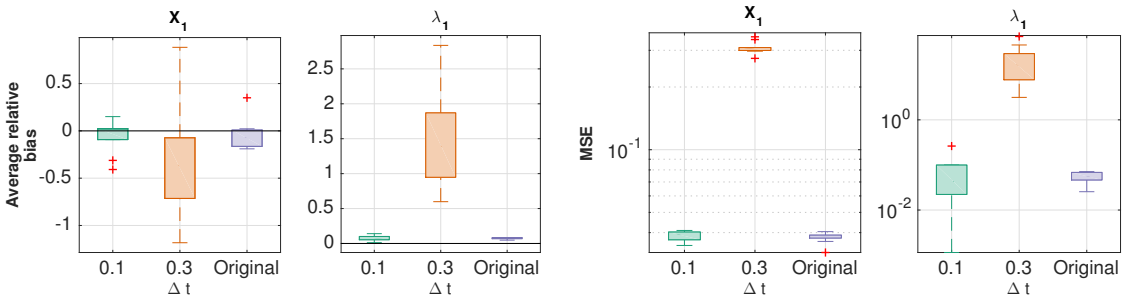

Figure 8. **Van der Pol system**: observation sample size. Boxplots of relative bias (*Left*) and MSE on a log scale (*Right*) obtained over 10 data sets for the three scenarios considered in Table 2.

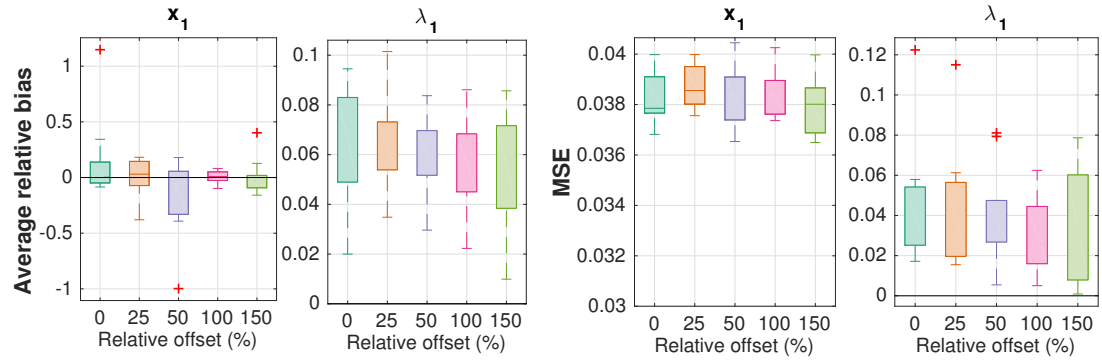

Figure 9. **Van der Pol system**: initialisation. Boxplots of relative bias, zoomed-in to a more informative scale (*Left*), and MSE(*Right*). Estimates obtained over 10 data sets with relative offsets $\{0\%, 25\%, 50\%, 100\%, 150\%\}$.

6 Discussion

Motivated by recent developments in mathematical biology, our study has focused on the accuracy of parameter inference in dynamical systems approximated by a UKF. We have selected three systems with features that are representative of many complex systems, including chaotic attractors and intrinsic stochasticity. The relative simplicity of these systems has allowed us to run hundreds of independent simulations. This has enabled us to obtain the distributions of two figures of merit, quantifying the distance from the known ground truth both in parameter and function space.

First, we looked at the dependence of the parameter estimation on the observation sample size, and we have done this by comparing the models with the original settings as implement in [7] with two different scenarios that contained just 30% of the original sample size. We have shown that for some models, the UKF requires a larger sample size (Lorenz attractor, Lotka-Volterra), or larger frequency (van der Pol) to produce reliable estimates, and this is specific to the particular dynamics involved (stochasticity, periodicity etc.).

Additionally, our particular focus has been on the dependence of the parameter estimation on the initialisation. We have shown that for an initialisation close to the ground truth the results of the seminal study in [7] can be reproduced. Nonetheless, even for perturbations in the order of 100% (a factor of 2), the results noticeably deteriorate. Our plots allow an objective quantification of this deterioration as a function of the initial deviation. The poorer performance for less informative initialisations suggests that the UKF is susceptible to local optima in the likelihood landscape. This appears to be an intrinsic feature of the sequential parameter updating scheme, whereby the relative weight of the corrector step, which updates the parameter distribution by assimilation of new data, exhibits a monotonic decrease proportional to the inverse of the number of data instances. This can render it difficult for the UKF to escape from a local attractor state in the vicinity of a poor initialisation.

In conclusion, our study provides a more cautionary picture than the original publication in [7] and suggests that a complementary fast preliminary parameter scanning scheme, as e.g. proposed in [8], appears to be indispensable to make the method applicable to complex systems for which little knowledge is available a priori.

Bibliography

[1] Baker S.M., Poskar C.H., and Junker, B.H. (2011) *Unscented Kalman filter with parameter identifiability analysis for the estimation of multiple parameters in kinetic models* EURASIP Journal on Bioinformatics and Systems Biology, **2011**:7

[2] Cohen J.E. (2004) *Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better* PLoS Biology, **2**(12)

[3] Gao, H., Li, W. G., Cai,L., Berry, C. and Luo, X. Y. (2015) *Parameter estimation in a Holzapfel-Ogden law for healthy myocardium* Journal of Engineering Mathematics **95**(1), 231–248

[4] Julier, S. J. and Uhlmann, J. K., (1997), *New extension of the Kalman filter to nonlinear systems.* AeroSense'97, 182–193.

[5] Macdonald, B., Higham, C., and Husmeier, D. (2015) *Controversy in mechanistic modelling with Gaussian processes* Journal of Machine Learning Research: Workshop and Conference Proceedings **37**, 1539–1547.

[6] Murphy, K.P. (2012) *Machine learning: a probabilistic perspective*, MIT Press, Cambridge.

[7] Sitz, A., Schwarz, U., Kurths, J. and Voss, H. U. (2002) *Estimation of parameters and unobserved components for nonlinear systems from noisy time series* , Phys. Rev. E, **66**(1), 016 – 210

[8] Strebel O. (2013) *A preprocessing method for parameter estimation in ordinary differential equations* Chaos, Solitons & Fractals 57 **93** –104

[9] Xia,J., Lamataa,J., Leeb,J., Moireauc,P., Chapellec,D. and Smith, N., (2011)*Myocardial transversely isotropic material parameter estimation from in-silico measurements based on a reduced-order unscented Kalman filter* Journal of Mechanical Behaviour of Biomechanical Materials **4**, 1090–1102